

ДОКЛАДЫ  
АКАДЕМИИ НАУК СССР

1988

ТОМ 299 № 2

ОТДЕЛЬНЫЙ ОТТИСК

В.Ю. ЛУНИН

**ВОССТАНОВЛЕНИЕ НЕДОСТАЮЩИХ СТРУКТУРНЫХ ФАКТОРОВ  
ПРИ РЕНТГЕНОСТРУКТУРНОМ ИССЛЕДОВАНИИ МАКРОМОЛЕКУЛ**

(Представлено академиком Б.К. Вайнштейном 13 XI 1986)

1. В процессе рентгеноструктурного исследования пространственной организации макромолекул при конечном разрешении  $d_{\min}$  находим функцию  $\rho(r)$ , являющуюся суммой отрезка ряда Фурье:

$$(1) \quad \rho(r) = \sum_{|s| \leq 1/d_{\min}} F(s) \exp\{i[\varphi(s) - 2\pi(s, r)]\}.$$

Будем называть эту функцию  $\rho(r)$  изображением (или, более полно, изображением функции распределения электронной плотности при разрешении  $d_{\min}$ ), чтобы подчеркнуть ее отличие от "истинного" распределения электронной плотности, отвечающего бесконечному ряду в (1). На практике точное определение изображения затруднено по двум причинам. Во-первых, фазы  $\varphi(s)$  структурных факторов определяются с ошибками (некоторые фазы могут быть вообще не определены), во-вторых, из-за конкретных условий рентгеновского эксперимента часть модулей  $F(s)$  структурных факторов может оказаться не измеренной.

Допустим, что нам известны модули  $F^0(s)$  структурных факторов с  $s \in S_m$  и фазы  $\varphi^0(s)$  с  $s \in S_p$ . Обозначим  $L$  множество изображений  $\rho(r)$  вида (1), у которых модули структурных факторов при  $s \in S_m$  и фазы при  $s \in S_p$  имеют значения  $F^0(s)$  и  $\varphi^0(s)$  соответственно, а прочие модули и фазы могут принимать произвольные значения. Наиболее распространенный подход к доопределению неизвестных структурных факторов (т.е. к выбору конкретного изображения из класса  $L$ ) — положить их равными нулю, или просто не включать в сумму (1). Более точные способы выбора функции из класса  $L$  основаны на наложении на функцию  $\rho(r)$  дополнительных требований, отражающих ее специфику. Например, неоднократно применялись ограничения на значения, принимаемые функцией  $\rho(r)$  вида [1—4]

$$\rho(r) \geq 0; \quad \rho_{\min} \leq \rho(r) \leq \rho_{\max}; \quad \rho(r) = 0 \text{ или } 1.$$

2. В этой работе предлагается способ более точного учета специфики области значений, принимаемых изображениями, отвечающими распределениям электронной плотности в кристаллах биологических макромолекул. Пусть  $\rho(r)$  рассчитана в узлах некоторой сетки в элементарной ячейке и  $\{\rho_j\}$  — совокупность ее значений в этих узлах. Построим гистограмму для этих значений т.е. разобъем интервал  $(\rho_{\min}, \rho_{\max})$  на  $K$  равных частей (бинов) и опреде-

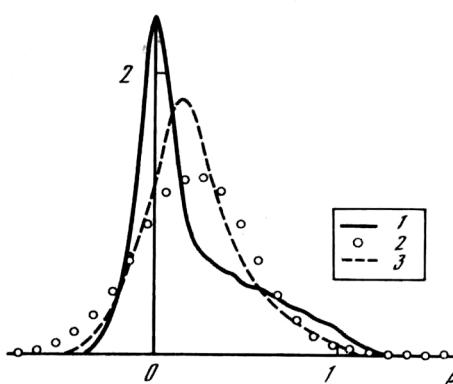


Рис. 1. Гистограммы для функций  $\rho_e$  (1),  $\rho_{sp}$  (2),  $\rho_f$  (3)

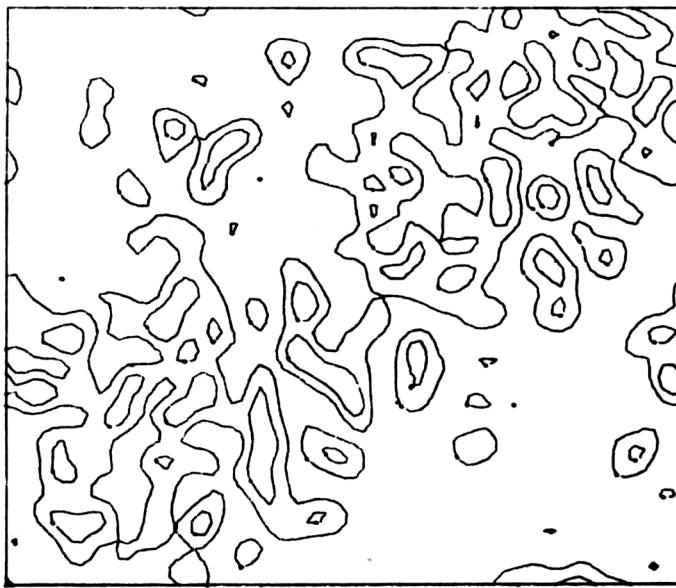


Рис. 2. Сечение  $z = 9/48$  функции  $\rho_e(r)$

лим частоты  $v_k$  попадания значений  $\rho_j$  в каждый из бинов. Анализ гистограмм для изображений белков с уже известной структурой позволяет сделать вывод, что они имеют характерные особенности, отличающие их от гистограмм случайно выбранных функций. Этот факт иллюстрирован на рис. 1, где приведены гистограммы для трех функций вида (1), отвечающих значению  $d_{\min} = 4 \text{ \AA}$ :

$\rho_e(r)$  определена по формуле (1), где в качестве  $F(s) \exp i\varphi(s)$  взяты точные структурные факторы, рассчитанные по атомной модели белка субтилизина;

$\rho_{sp}(r)$  — в качестве  $F(s)$  в (1) взяты точные значения, рассчитанные по модели субтилизина, а  $\varphi(s)$  определены датчиком случайных чисел;

$\rho_r(r)$  — из синтеза  $\rho_e(r)$  исключено около 18% рефлексов центральной зоны.

Рис. 1 иллюстрирует также тот факт, что часто применяемое требование неотрицательности функции  $\rho(r)$  носит приближенный характер. Даже при абсолютно точном определении модулей и фаз структурных факторов изображение (1) при среднем разрешении будет иметь области, в которых  $\rho(r) < 0$ , причем объем таких областей будет почти тем же, что и в случае случайного выбора фаз.

Для разных белков и разного разрешения гистограммы для точных изображений  $\rho_e(r)$  могут несколько различаться и определение эталонной гистограммы  $\{v_k^0\}$  для еще неизвестного объекта представляет отдельную задачу. Мы ограничимся в этой работе важным частным случаем, когда эталонную гистограмму можно считать известной. Это случай, когда известна структура гомологичного белка, гистограмма изображения которого может быть взята в качестве эталона.

3. Допустим теперь, что нам известна для исследуемого объекта эталонная гистограмма  $\{v_k^0\}$ . Применим эту информацию для выбора функции из введенного выше класса  $L$ . Пусть  $\rho^c(r)$  некоторая функция из этого класса. Мы можем рассчитать гистограмму  $\{v_k^c\}$  для этой функции и сравнить ее с эталонной:

$$(2) \quad Q(\rho^c) = \frac{1}{K} \sum_{k=1}^K \frac{(v_k^c - v_k^0)^2}{v_k^0} .$$

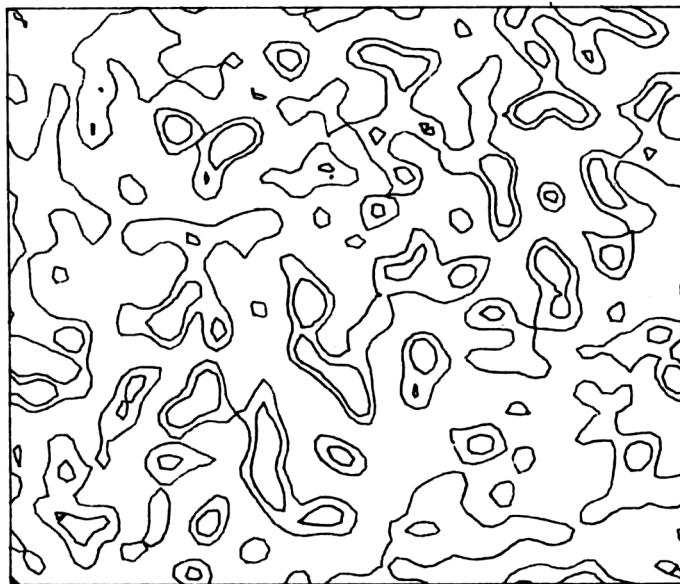


Рис. 3. Результат исключения из синтеза  $\rho_e(\mathbf{r})$  18% рефлексов

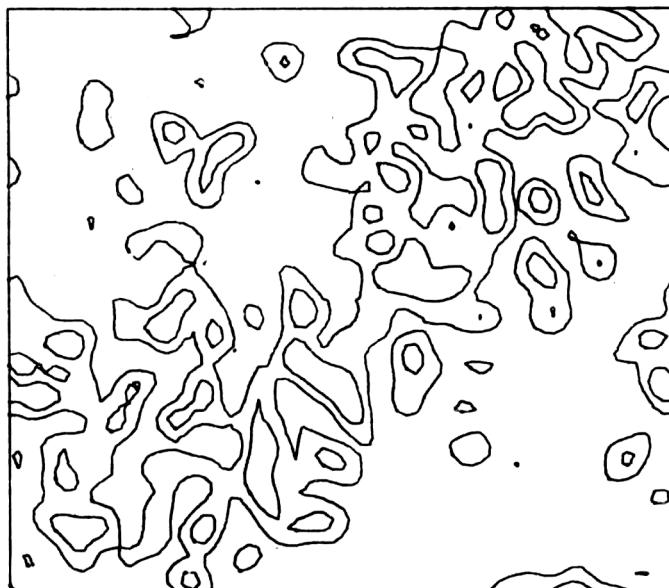


Рис. 4. "Восстановленное" изображение

Сформулируем теперь критерий отбора функции  $\rho(\mathbf{r})$  из класса  $L$  следующим образом:

**Критерий отбора.** Среди функций  $\rho(\mathbf{r})$  класса  $L$  найти ту, для которой величина  $Q(\rho)$  минимальна.

Это означает, что мы разрешаем неизвестным структурным факторам принимать любые (не обязательно нулевые значения), но требуем при этом, чтобы

гистограмма получаемой функции  $\rho(r)$  была по возможности наиболее близка к эталонной.

4. Полезность сформулированного критерия для восстановления значений структурных факторов, модули которых не были измерены, демонстрирует следующий тест. В качестве тестового объекта взята атомная модель белка субтилизина и помещена в элементарную ячейку размерами  $73 \times 64 \times 48 \text{ \AA}$  в пространственной группе  $P2_12_12_1$ . По координатам атомов рассчитаны структурные факторы и построен синтез разрешения  $4 \text{ \AA}$  (рис. 2). По этому синтезу определены эталонные частоты. Для того, чтобы проверить, насколько гистограмма чувствительна к изменению ориентации молекулы, проведен аналогичный расчет для молекулы субтилизина в другой ориентации. Введенный аналогично (2) критерий соответствия принял значение  $0,1 \cdot 10^{-4}$ , что определило предел, до которого имеет смысл опускать значение минимизируемого критерия при выборе изображения из класса  $L$ .

Далее была имитирована ситуация отсутствия информации о части модулей структурных факторов. Около 18% структурных факторов объявлены неизвестными; поставлена задача их определения путем минимизации критерия типа (2). В качестве стартовых значений для неизвестных структурных факторов взяты нулевые значения, что дало изображение, показанное на рис. 3. Далее была проведена минимизация при помощи специально созданной программы. В результате 5 шагов спуска по антиградиенту значение критерия снизилось с  $0,3 \cdot 10^{-2}$  до  $0,1 \cdot 10^{-4}$ . Изображение  $\rho(r)$ , рассчитанное с использованием определенных в процессе минимизации структурных факторов, приведено на рис. 4. Точность определения неизвестных модулей структурных факторов составило

$$\sum |F^o - F^c| / \sum F^o = 0,50,$$

где  $F^o$  — точные значения модулей,  $F^c$  — значения, полученные в результате минимизации; сумма взята по всем 352 рефлексам, считавшимся неизвестными. Средняя ошибка определения фаз по 170 нецентросимметричным рефлексам составила  $36^\circ$ , из 182 центросимметричных рефлексов неверно определен знак у 34 рефлексов.

Научно-исследовательский вычислительный центр  
Академии наук СССР,  
Пущино Московской обл.

Поступило  
26 XI 1986

#### ЛИТЕРАТУРА

1. Сирота М.И., Симонов В.И. — Кристаллография, 1970, т. 15, № 4, с. 681—685. 2. Вайнштейн Б.К., Хачатуян А.Г. — Там же, 1977, т. 22, № 4, с. 706—710. 3. Qurashi M.M. — Acta cryst., 1953, vol. 6, № 1, p. 106. 4. Davies A.R., Rollett J.S. — Ibid., 1976, vol. A32, № 1, p. 17—23.